Cheat sheet for model uncertainty assessment

Novartis Contributors: Andrew M Stein, Jeff D Kearns, Jaeyeon Kim, Alison Margolskee

Purpose of this document

The Pedigree Table on the back of this document helps a modeler or reviewer assess the uncertainty of a model's predictions in a structured way [1, 2]. This tool should be applied during scoping, to ensure effort is placed on the most critical aspects of the model, and then applied again at the end of the activity, to ensure the model is fit for the purpose of informing the decision. In general, the more important the decision, the higher the scores should be. Because model uncertainty is difficult to rigorously quantify, this assessment is qualitative, and reviewers may differ in their assessment. This difference in opinion can be useful information and it is recommended that the results from different reviewers be shown simultaneously to highlight areas of disagreement [2]. Figure 1 shows one way to visualize the results of the uncertainty assessment. This figure is illustrative and does not refer to specific models in the literature.

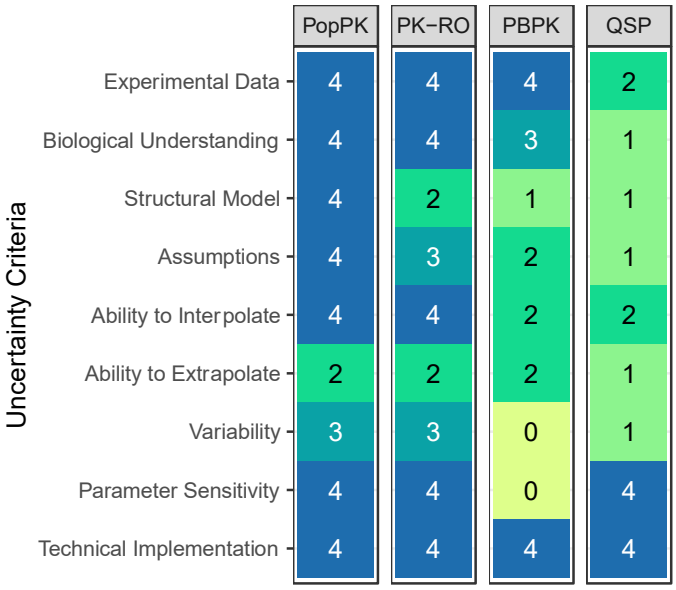

Figure 1 An example pedigree table for four hypothetical examples: PopPK, PK-RO (i.e. PKPD receptor occupancy), PBPK, and QSP. Scores range from 0 (very poor assessment of uncertainty) to 4 (very good assessment of uncertainty)

Important Considerations

Fit for Purpose: What is the purpose of this model? What are the key behaviors that it must capture? While complex models may be initially employed for hypothesis exploration, use the simplest model that addresses the specific question; reevaluate the model if the question changes.

Consequences of incorrect prediction: If the model prediction is wrong, what are the consequences? What level of uncertainty can be tolerated? The pedigree table scores should reflect this.

Range of predictions: How does the model uncertainty affect the decision or recommendation? Have you clearly defined a range of model predictions (i.e. 'best', 'median', 'worst' cases)? This representation is often more accessible to non-modelers versus formal uncertainty metrics.

References

1. Saltelli, A. "A short comment on statistical versus mathematical modelling." *Nature communications* 10.1 (2019), 1
2. Van Der Sluijs, JP., et al. "Combining quantitative and qualitative measures of uncertainty in model-based environmental assessment: the NUSAP system." Risk Analysis: An International Journal 25.2 (2005): 481.

Pedigree table for model uncertainty assessment

Criteria	Very Poor (0)	Poor (1)	Fair (2)	Good (3)	Very Good (4)
What is the quality of the experimental data?	No data	Very limited data, in some cases, educated guesses inform the model	Limited data, often based on old experiments using unvalidated assays	Enough trusted data for good description of most critical aspects of system	All data from validated assays and reproducible experiments. Data fully describes system
Is there consensus on biological mechanisms included in the model and their mathematical description?	Crude speculation	Embryonic field	Limited consensus	Accepted theory	Well-established theory
How complete was the exploration of structural model?	Only a single structural model considered, without clear rationale	A few structural models were considered	A few structural models representing key sources of uncertainty	Many structural models, including random model perturbations were used	Structural models representing all known sources of uncertainty were used in predictions
Have you clearly articulated the assumptions and their consequences?	Assumptions are not clearly listed	Assumptions are listed, but impact of many assumptions not explored or explained	Many assumptions. They are clearly explained. Impact was explored but is difficult to understand	Moderate number of assumptions, they are clearly listed, and impact of assumptions is clear	Few assumptions, they are clearly listed, and impact of assumptions is clear
Can your model accurately interpolate?	Not designed for interpolation		Model fits are consistent with data		Model fits are validated using standard diagnostics (VPCs, residuals)
Can your model accurately extrapolate?	Not designed for extrapolation	Weak confidence in model ability to extrapolate	Moderate confidence in ability to extrapolate	Strong confidence in ability to extrapolate	Many surprising extrapolated predictions in system of interest that were confirmed
Have you addressed Intra- and Inter-subject variability?	Variability is not included in the model.	Unknown where variability should be placed, so many guesses were made		Reasonable hypotheses exist for main sources of variability and they are included	The key sources of variability are well understood and included
Have you explored parameter sensitivity and identifiability?	No	Local analysis only	Global analysis, over a small (~10%) parameter range	Global analysis over a large (~10-fold) parameter range	Global analysis over well-informed, sufficiently large range or established theory exists for understanding system
Have you performed quality control on your model?	Used an unvalidated computing environment, one modeler involved, no audit		Used a validated reproducible, computing environment, modeler performed self-audit		Used a validated, reproducible computing environment with a careful audit of the most likely sources of errors in programming by an independent modeler